

C a d e r n o s T é c n i c o s

Introdução à preservação de acervos digitais



Secretaria de Estado de Cultura
Arquivo Público Mineiro

Cadernos Técnicos

Introdução à Preservação de Acervos Digitais

Belo Horizonte
Secretaria de Estado de Cultura
Arquivo Público Mineiro
2005

Amorim, Eliane Dutra

Introdução à preservação de acervos digitais. / Eliane Dutra Amorim, Carlos Eduardo Rodrigues Lopes, Eduardo Alves do Valle Júnior. – Belo Horizonte: Secretaria de Estado de Cultura / Arquivo Público Mineiro, 2005.

42p. (Cadernos Técnicos)

ISBN 85-99528-01-7

1. Acervos digitais – preservação. I. Lopes, Carlos Eduardo Rodrigues. II. Valle Júnior, Eduardo Alves do. III. Arquivo Público Mineiro. IV. Título. V. Série.

FICHA TÉCNICA

Aécio Neves da Cunha
Governador do Estado de Minas Gerais

Maria Eleonora Barroso Santa Rosa
Secretária de Estado de Cultura

Marcelo Braga de Freitas
Secretário-Adjunto de Estado de Cultura

Renato Pinto Venâncio
Superintendente do Arquivo Público Mineiro

Alice Oliveira de Siqueira
Diretora de Acesso à Informação e Pesquisa

Augusta Aparecida Cordoval Caetano
Diretora de Gestão de Documentos

Emerson Nogueira Santana
Diretor de Arquivos Permanentes

Pedro de Brito Soares
Diretor de Conservação de Documentos

Equipe Técnica

Coordenação
Eliane Dutra Amorim

Elaboração de Textos
Carlos Eduardo Rodrigues Lopes
Eliane Dutra Amorim

Consultoria e Revisão Técnica
Eduardo Alves do Valle Jr.

1. INTRODUÇÃO.....	7
2. O QUE É PRESERVAÇÃO DIGITAL?	8
2.1. DESAFIOS DA PRESERVAÇÃO DIGITAL	9
2.2. POR QUANTO TEMPO PRESERVAR?	9
3. PRESERVAÇÃO DIGITAL VS. TRADICIONAL.....	12
3.1. MEIOS DE ARMAZENAMENTO	12
3.2. DEGRADAÇÃO E DETERIORAÇÃO – VISÍVEL VS. INVISÍVEL	13
3.3. RESPONSABILIDADE PELA PRESERVAÇÃO.....	13
4. METADADOS	14
4.1. O QUE SÃO METADADOS?.....	14
4.2. CATEGORIAS E FUNÇÕES	14
4.3. POR QUE OS METADADOS SÃO IMPORTANTES?	15
4.4. PADRÕES DE METADADOS.....	16
4.5. METADADOS DE PRESERVAÇÃO.....	16
4.6. ARMAZENANDO METADADOS.....	17
5. ARQUIVOS DIGITAIS DE PRESERVAÇÃO	19
5.1. FORMATOS PARA DOCUMENTOS VISUAIS.....	20
5.1.1. <i>Tagged Image File Format – TIFF</i>	20
5.1.2. <i>Joint Photographic Experts Group – JPEG</i>	21
5.2. FORMATOS PARA DOCUMENTOS TEXTUAIS	22
5.2.1. <i>Texto Puro</i>	23
5.2.1.1 ASCII	24
5.2.1.2 Unicode	24
5.2.2. <i>eXtensible Markup Language – XML</i>	24
5.2.3. <i>Portable Document Format – PDF</i>	25
5.3. PADRONIZAÇÃO DE FORMATOS.....	26
5.4. TABELA COMPARATIVA.....	27
6. MÍDIAS DE ARMAZENAMENTO.....	28
6.1. DISCOS RÍGIDOS.....	29
6.2. FITAS MAGNÉTICAS	30
6.3. DISCOS ÓTICOS.....	32
6.4. COMPARAÇÃO ENTRE AS MÍDIAS DE ARMAZENAMENTO.....	35
6.5. AMBIENTE DE ARMAZENAMENTO	35
7. ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL.....	37
7.1. REFRESCAMENTO.....	37
7.2. MIGRAÇÃO	38
7.2.1. <i>Relação entre Refresco e Migração</i>	39
7.3. EMULAÇÃO.....	40
8. CONCLUSÃO.....	42
9. REFERÊNCIAS.....	43

1. INTRODUÇÃO

A partir de meados da década de 90, diversos órgãos brasileiros de documentação passaram a utilizar a tecnologia digital como ferramenta para otimizar as ações de preservação e acesso. Esta tendência vem crescendo de forma significativa e, nos últimos anos, os projetos visando a digitalização de acervos documentais passaram a consumir parte significativa dos recursos liberados pelas agências de financiamento.

No entanto, os investimentos e esforços para geração destas coleções digitais correm sérios riscos. A maior parte dos registros digitais, criados ao longo dos anos, pode se tornar inacessível, a curto prazo, devido à obsolescência tecnológica, à fragilidade dos meios de armazenamento e, principalmente, à ausência de estratégias de preservação digital.

As estratégias de preservação digital, neste sentido, precisam ser consideradas a base dos projetos voltados à digitalização de acervos; caso contrário, os registros digitais, que vêm sendo produzidos, podem ser perdidos - parcial ou totalmente - entre cinco e dez anos.

O Arquivo Público Mineiro (APM), visando a preservação dos registros digitais, gerados a partir de projetos de digitalização de seu acervo, constituiu um grupo de estudos, com o objetivo de desenvolver um plano de preservação digital para a instituição.

Durante o ano de 2004, o grupo formado por técnicos do APM, assim como por estudantes do Núcleo de Processamento Digital de Imagens (NPDI) da Universidade Federal de Minas Gerais, sistematizou informações sobre preservação digital para subsidiar estes trabalhos. Partes destas informações foram reunidas nesta publicação, que tem por objetivo divulgar noções básicas sobre preservação digital e, assim, auxiliar e conscientizar as instituições sobre a importância de ações voltadas à preservação digital, em todas as etapas de execução de projetos envolvendo digitalização de acervos.

Destaca-se, entretanto, que o foco desta publicação é a preservação de registros gerados a partir da conversão de documentos analógicos para o meio digital; apesar de alguns tópicos poderem auxiliar a preservação de documentos produzidos originalmente em meio digital.

2. O QUE É PRESERVAÇÃO DIGITAL?

A tecnologia digital, como forma de otimizar ações de preservação e acesso, vem sendo adotada por instituições do mundo inteiro devido às inúmeras vantagens que oferece. O Brasil segue também essa tendência, e, recentemente, projetos voltados para digitalização de acervos documentais têm se beneficiado de significativos financiamentos oferecidos por agências governamentais.

Nesse contexto, a preservação digital deve ser entendida sob dois aspectos. Primeiro a conversão dos documentos analógicos, gerando cópias digitais que permitam o acesso à informação e evitem o manuseio direto e constante dos originais. Segundo, a garantia do acesso, pelas gerações atuais e futuras, aos registros digitais, impedindo que a obsolescência tecnológica e a fragilidade dos meios de armazenamento tornem esses registros ilegíveis.

Projetos que negligenciem a preservação, a longo prazo, podem tornar os registros irremediavelmente inacessíveis em menos de uma década. A perda do registro digital, além do desperdício de recursos, ameaça o acervo como um todo — uma vez que o documento original deverá ser submetido à nova digitalização (com resultantes riscos e desgastes). Se o original houver envelhecido, danificado ou desaparecido, a perda da cópia torna-se crítica.

A fim de garantir o acesso e ampliar a vida útil dos registros digitais, é necessária a adoção de procedimentos de preservação em todo ciclo de vida do registro, ou seja, do ato de sua criação à destinação final (guarda permanente ou eliminação). Os procedimentos implantados devem contemplar:

- Gerência de metadados;
- Definição dos formatos de arquivos;
- Seleção de mídias de armazenamento;
- Controle do ambiente de guarda do acervo;
- Definição de critérios para manuseio do acervo;
- Seleção de técnicas de preservação ativa do acervo digital, como a migração, refrescamento e emulação.

Dada a importância destas ações, detalhes sobre cada uma serão apresentados nas seções que compõem este guia.

2.1. Desafios da Preservação Digital

Os meios de armazenamento digital possuem vida útil inferior aos suportes tradicionais (papel, microfilme, pergaminhos). Este período varia, ainda, em decorrência da qualidade, do manuseio e das condições de guarda do suporte.

A vida física dos meios de armazenamento é um aspecto importante da preservação digital, entretanto, garantir a durabilidade das mídias não significa garantir o acesso aos dados. A legibilidade dos dados digitais está associada à existência de programas e equipamentos capazes de interpretar as informações contidas nos suportes. Tem efeito nulo a preservação de suportes por uma, duas ou mais décadas se os interpretadores deixarem de existir. A evolução tecnológica é veloz, e, após períodos muitas vezes inferiores a cinco anos, programas e equipamentos tornam-se obsoletos, resultando, conseqüentemente, na ilegibilidade dos dados.

A dependência de equipamentos e programas, a fragilidade dos meios de armazenamento e a obsolescência tecnológica são os maiores desafios para preservação dos registros digitais.

Os exemplos a seguir ilustram questões nas quais a degradação das mídias ou a obsolescência tecnológica comprometeu o acesso às informações digitais [CPA 96]:

- A Agência Espacial Americana — NASA — perdeu cerca de 20% dos dados coletados sobre Marte em 1976 em virtude de degradação/deterioração das fitas magnéticas.
- Em 1975, apenas dois computadores no mundo permitiam o acesso ao censo americano de 1960.
- Os dados sobre o uso de terras no Estado de Nova Iorque levantados em 1960 tornaram-se inacessíveis após vinte anos porque já não existiam os programas necessários para sua leitura.

2.2. Por Quanto Tempo Preservar?

Papéis ácidos, fotografias, filmes, registros analógicos em geral, são vulneráveis ao ambiente que os cerca, porém, a adoção de medidas de preservação dez, vinte, ou cinquenta anos, após sua criação, ainda permite a recuperação parcial ou total

dos dados. Tratando-se de registros digitais, entretanto, as fragilidades dos suportes, associadas à obsolescência tecnológica, levam à perda total da informação em poucos anos. Neste cenário, é comum a divisão dos esforços de preservação digital em três categorias [HOWELL 04]:

- Preservação em curto prazo;
- Preservação em médio prazo;
- Preservação em longo prazo.

A Tabela 1 caracteriza cada um destes tipos de preservação.

PRESERVAÇÃO			
	CURTO PRAZO	MÉDIO PRAZO	LONGO PRAZO
Definição	Acesso enquanto a informação e a tecnologia forem correntes.	Acesso por um determinado período de tempo, porém não indefinidamente.	Acesso indefinidamente.
Objetivos	Manter a usabilidade.	Preservar a usabilidade, o contexto e o conteúdo da informação não necessariamente em uma cópia digital exata, ou com apresentação original.	Preservar a usabilidade, o contexto e o conteúdo da informação não necessariamente em uma cópia digital exata, ou com apresentação original.
Plataforma computacional	Plataforma original.	Mudança necessária quando a plataforma original tornar-se obsoleta.	Mudança necessária quando a plataforma original tornar-se obsoleta.
Formatos de texto e imagem	Formato original.	Formato original, ou formatos que sejam padrões estabelecidos.	Formatos que sejam padrões estabelecidos.
Atualização	Nenhuma.	Planejada.	Planejada.
Armazenamento	Portador do registro original.	Possivelmente no portador do registro original.	Suporte padrão de mercado, de alta capacidade e com comprovada longevidade.

PRESERVAÇÃO

	CURTO PRAZO	MÉDIO PRAZO	LONGO PRAZO
Local de armazenamento	Instituição produtora.	Instituição produtora e se possível uma instituição externa.	Instituição produtora e uma instituição externa.
Estratégias de preservação	Cuidados básicos.	Refrescamento. Migração.	Refrescamento. Migração. Emulação.
Frequência de cuidados	Quando necessário.	Planejado considerando o armazenamento e as estratégias de preservação do acervo.	Planejado considerando o armazenamento e as estratégias de preservação do acervo.
Descarte	Quando a tecnologia está obsoleta.	Planejado.	Planejado.

Tabela 1: Preservação de curto, médio e longo prazo.**FONTE – [HOWELL 04]**

3. PRESERVAÇÃO DIGITAL VS. TRADICIONAL

3.1. Meios de Armazenamento

Os registros digitais, assim como o papel, a fotografia, os filmes e outros suportes tradicionais, estão sujeitos à degradação e deterioração causadas por agentes externos e internos.

Agentes internos estão relacionados à composição do material constituinte do suporte. Os agentes externos, por sua vez, estão relacionados às condições de uso e do ambiente no qual os suportes se encontram. São agentes externos causadores de danos, dentre outros:

- Temperatura;
- Umidade;
- Poluição;
- Fogo;
- Água;
- Manuseio incorreto do suporte.

Para impedir ou retardar os efeitos destes agentes, devem ser adotadas medidas de controle ambiental da área de guarda, cuidados com manuseio e prevenção de desastres (inundações, incêndios, sabotagem). Estas medidas são utilizadas para preservação dos registros digitais e tradicionais, no entanto, os registros digitais são fisicamente mais vulneráveis.

Um aspecto importante com relação aos registros digitais está em sua alta capacidade de armazenamento de informações. O impacto da perda de uma mídia digital é, em termos quantitativos, superior ao da perda de um registro tradicional.

Por outro lado, uma das grandes vantagens do registro digital, é a possibilidade da redundância. Como a informação digital é perfeitamente duplicável, é possível manter uma cópia exata do acervo em um segundo depósito. Dada a fragilidade dos suportes e sua vulnerabilidade em caso de sinistro, essa redundância é altamente recomendável.

3.2. Degradação e Deterioração – Visível vs. Invisível

Uma das diferenças entre a preservação digital e a tradicional está na visibilidade dos estragos causados pela degradação e deterioração nos registros. Quando um registro tradicional, por exemplo, um livro, fotografia ou mapa, apresenta problemas em virtude da umidade, constata-se a presença de fungos, e o manuseio incorreto leva à presença de rasgos e dobras. Resumindo, independente do agente causador, os danos são, em grande parte, perceptíveis a olho nu.

Nos registros digitais, entretanto, a ação de campos magnéticos, umidade excessiva, dentre outras possíveis causas de degradação e deterioração, são imperceptíveis aos olhos humanos. Esta característica representa um risco adicional à já frágil constituição dos suportes digitais, pois o problema acaba sendo constatado somente quando os danos são irreversíveis.

3.3. Responsabilidade pela Preservação

Responsabilidades sobre as medidas de preservação para registros tradicionais concentram-se, principalmente, em arquivos, bibliotecas e museus. Já com os registros digitais, que requerem medidas de preservação durante todo seu ciclo de vida, a rede de responsabilidade se expande, sendo formada por todos os agentes que acompanham o registro, desde o produtor até os órgãos de documentação (arquivos, bibliotecas e museus).

Observa-se, ainda, que as ações de preservação de registros digitais estão interligadas e o sucesso de um programa de preservação está relacionado à participação e cooperação de todos os agentes.

4. METADADOS

4.1. O que são Metadados?

Metadados são, literalmente, “dados sobre os dados”, entretanto, podem ser melhor definidos como a **informação estruturada** que descreve um registro [HUTT 04], contemplando informações administrativas, estruturais, de classificação de conteúdo, proveniência, relacionamento entre documentos, dentre outras.

A informação contida em um metadado pode ser explícita como o autor, a data de criação, a descrição de uma imagem. Ou ser mais complexa, tais como informações técnicas a respeito do dispositivo de captura (câmara digital ou scanner), formatos de arquivo utilizados e resolução da imagem digital [TASI 02c].

4.2. Categorias e Funções

São várias as abordagens e linhas de pensamento a respeito da categorização de metadados. Existe, no entanto, um consenso sobre funções básicas que eles devem desempenhar. A Tabela 2 apresenta uma divisão simplificada e esclarecedora que atende às necessidades deste guia.

Categoria	Função	Exemplo
Descritivo	Fornecer identificação, localização e acesso aos registros.	Catálogo bibliográfico (autor, título etc.), descrição arquivística (proveniência, conteúdo, datas-limite etc.), vocabulários controlados (lista de termos utilizados como ponto de acesso).
Administrativo Técnico Preservação	Fornecer detalhes técnicos sobre a criação e armazenamento de registros.	Formato de arquivo, <i>software/hardware</i> e suporte utilizados, migração do formato x para y.
Estrutural	Fornecer informações sobre a ordem lógica e relação entre registros.	Seqüência das imagens em um álbum fotográfico ou entre páginas e capítulos em um livro.
Direitos Autorais	Fornecer informações sobre os aspectos legais de uso e reprodução de registros.	Legislação de direito autoral referente ao registro.

Tabela 2: Categorias de metadados.

Fonte – [HUTT 04]

4.3. Por que os Metadados são Importantes?

O uso de metadados em organizações, arquivos e museus fornece aos usuários toda a informação relevante acerca de um documento. São potencializadas pelo uso de metadados as seguintes ações envolvendo registros digitais [GRAY 04]:

- Acesso: metadados ricos e consistentes tornam a busca por informações mais rápida e eficiente.
- Classificação: metadados ajudam na classificação e categorização de registros de acordo com um ou mais vocabulários.
- Pesquisa: metadados tornam a recuperação da informação não apenas mais rápida e fácil, mas também mais precisa e barata.
- Preservação: metadados técnicos sobre a criação de registros digitais asseguram que estes sejam acessíveis no futuro.

- Gerência de Registros: descrição padronizada de registros reduz os custos referentes à organização e guarda, além de incrementar a interoperabilidade das informações.
- Gerência de Direitos de Reprodução: direitos de propriedade intelectual e condições de uso e reprodução de um registro podem ser capturados por metadados.

4.4. Padrões de Metadados

A estrutura do metadado é composta por um conjunto de elementos que, organizados, orientam a descrição de um registro. Esses elementos são de livre escolha; no entanto, existem esforços para a criação de padrões com o objetivo de facilitar o intercâmbio de informações.

O desenvolvimento de padrões de metadados não ocorreu de forma linear. Ao longo das últimas décadas, diferentes grupos têm desenvolvido, de forma independente, métodos de descrição, organização e gerência de informações para seus conjuntos de registros. Desta forma, diversos padrões de metadados encontram-se à disposição de arquivistas. A Tabela 3 apresenta alguns dos padrões de metadados mais populares atualmente.

Padrão	Função
Dublin Core	Dados sobre páginas da WEB.
Government Information Locator Service (GILS)	Informações governamentais.
Consortium for the Interchange of Museum Information (CIMI)	Informações sobre museus.
Machine Readable Cataloging (MARC)	Catálogo bibliográfico.
Encoded Archival Description (EAD)	Descrição arquivística.

Tabela 3: Padrões de metadados.

4.5. Metadados de Preservação

Metadados de preservação são definidos como o conjunto de informações técnicas necessárias para garantir o acesso e a preservação dos registros digitais ao longo do tempo [NLA 03].

Informações contidas nos metadados de preservação ajudam a enfrentar os desafios da obsolescência e possibilitam a interpretação dos registros digitais não

obstante as mudanças tecnológicas, além de auxiliar no planejamento e execução das três estratégias primárias de preservação para recursos digitais: refreshamento, migração, e emulação (ver Capítulo 7).

Estas características, portanto, transformam os metadados de preservação em itens fundamentais em qualquer estratégia de preservação.

4.6. Armazenando Metadados

São duas as formas de se armazenar os metadados referentes a um registro digital:

- As informações são inseridas no próprio registro. É exemplo desta abordagem o uso de áreas específicas do cabeçalho de alguns formatos de arquivo tais como JPEG, TIFF e DOC (Microsoft Word).
- Os metadados não se encontram fisicamente atrelados aos registros sendo que, em muitas situações, metadados e registros estão armazenados em suportes distintos. Esta abordagem implica na criação de bases de dados nas quais os metadados e os registros são relacionados. O catálogo de uma biblioteca, contendo informações sobre título, autor, editora, é um exemplo desta forma de armazenamento.

A melhor forma de armazenamento de metadados é determinada em virtude de diversos aspectos: gerência, tipo (texto, imagem, som) e necessidades de uso dos registros e seus metadados.

Metadados inseridos nos registros formam uma unidade que não corre o risco de ser separada. No entanto, o custo de gerência desta informação torna-se mais alto já que alterações nos metadados implicam em alterações nos registros. Um aspecto negativo de metadados embutidos em registros está na dificuldade do intercâmbio de informações devido, principalmente, à inexistência de padrões para o armazenamento de metadados. Por exemplo, metadados armazenados em um arquivo JPEG não obedecem à mesma estrutura dos metadados armazenados em um arquivo TIFF. A inserção de metadados além dos suportados pelo formato¹ pode ainda causar incompatibilidade entre o registro e o interpretador do formato, podendo inviabilizar o acesso ao documento (ver Seção 5.1.1).

Estando em uma base de dados, as informações dos metadados são mais facilmente recuperadas e apresentam diversas formas de apresentação. O grande

¹ Existem formatos que possibilitam a inserção de novos campos em seu cabeçalho.

desafio desta abordagem, entretanto, está na manutenção da base. Faz-se necessária a limitação do acesso, a fim de evitar alterações não autorizadas, e a manutenção da relação entre o registro e o metadado, através de um identificador comum. A má administração da base de dados introduz o risco da total perda dos metadados.

5. ARQUIVOS DIGITAIS DE PRESERVAÇÃO

A correta escolha do formato de arquivo digital a ser utilizado torna mais fácil a tarefa de preservação de documentos digitais. Um arquivo digital de preservação deve levar em consideração a acessibilidade e a utilidade do documento, a longo prazo (próximas décadas). Formatos não apropriados podem resultar em dificuldades futuras em questões como legibilidade, capacidade de migração do documento, dentre outras.

Algumas recomendações devem ser seguidas no momento da escolha do formato para o documento digital, seja este visual (imagens, gráficos) ou textual:

- Dê preferência a formatos não-proprietários, ou seja, formatos cuja especificação seja pública ou cujo desenvolvimento não esteja ligado a uma empresa específica [TASI 02a]. O uso de formatos proprietários torna às vezes impossível obter uma especificação oficial, além de introduzir o risco da descontinuidade do formato pela empresa criadora. Em optando por um formato proprietário, prefira aqueles amplamente utilizados e documentados.
- Utilize sempre formatos que sejam padrões *de facto*.² Uma vez que mais organizações, empresas e pessoas utilizam um determinado formato, sua vida útil é estendida, bem como o suporte que ele desfruta.
- Evite a compressão dos dados. Sendo a compressão necessária, utilize formatos com compressão sem perdas ou privilegie a qualidade ao máximo em detrimento à redução do tamanho final do arquivo. Isso deve ser feito considerando-se as necessidades futuras, ainda não previstas, da informação.
- Evite a criptografia dos dados. Se for necessário fazer uma migração de formato posteriormente, a criptografia pode introduzir dificuldades muitas vezes intransponíveis.

Embora tais recomendações sejam válidas, nota-se que poucos são os formatos que conseguem estar de acordo com todas, em especial quando se tratam de documentos visuais. O formato TIFF para documentos visuais, por exemplo, mesmo sendo um padrão industrial *de facto* e geralmente criado sem compressão, é um formato proprietário licenciado pela *Adobe Systems Incorporated*. Dentre os formatos que seguem todas as recomendações está o XML, formato de documento textual aberto, padrão industrial e que não utiliza nenhuma compressão.

² Padrões *de facto* são aqueles estabelecidos no meio comercial e industrial através do uso generalizado ou pelo domínio de mercado.

Apresentam-se nas seções que seguem maiores informações sobre os principais formatos de arquivos digitais de preservação. Inicialmente consideram-se documentos visuais. São dois os formatos listados: TIFF e JPEG. Para documentos textuais consideram-se os formatos XML e PDF, além das codificações de caracteres ASCII e Unicode.

5.1. Formatos para Documentos Visuais

Documentos visuais englobam todo tipo de documento cujo conteúdo visual seja de interesse para preservação. São as chamadas **imagens digitais** e estão incluídas nesta classe de documentos:

- Fotografias, mapas, desenhos, plantas arquitetônicas digitalizadas a partir de originais em papel, microfilme ou negativo;
- Imagens criadas digitalmente utilizando *softwares* de criação e edição de imagens;
- Fotografias produzidas por câmeras fotográficas digitais;

5.1.1. Tagged Image File Format – TIFF

O formato TIFF é atualmente o padrão para a criação de arquivos de preservação. O TIFF foi desenvolvido pela *Aldus Corporation* conjuntamente com a *Microsoft Corporation* em 1986, ficando os direitos de cópia sob a guarda da *Aldus*, que veio a se fundir com a *Adobe*. Conseqüentemente, a *Adobe* possui, hoje, os direitos de cópia das especificações do formato que se encontra na versão 6.0 [ADOBE 92].

Várias são as características que tornam o formato TIFF a melhor escolha para a preservação de documentos visuais.

- Desde o início do seu desenvolvimento, o formato foi projetado de forma a ser completamente independente da plataforma utilizada. Um arquivo TIFF tem o mesmo comportamento nos mais diversos sistemas, seja ele Windows, Linux ou Macintosh.
- Como consequência desta independência, o TIFF é suportado por várias aplicações de processamento de imagens digitais.
- O TIFF tornou-se um padrão industrial *de facto* mesmo sendo um formato proprietário.
- Imagens digitais podem ser armazenadas com ou sem compressão. Dentre os métodos de compressão suportados estão o LZW e o JPEG que, para fins de preservação, devem ser evitados (especialmente o JPEG sobre TIFF).

- A grande flexibilidade do formato, novos campos descritivos — as chamadas *tags* — podem ser inseridos no cabeçalho do formato.

Esta flexibilidade, entretanto, deve ser utilizada com cuidado. *Tags* que não sejam aprovadas pela *Adobe* e descritas na especificação do formato podem ser incompatíveis com alguns aplicativos (mesmo *tags* presentes na descrição podem ser incompatíveis com algumas aplicações — esse recurso deve ser usado com parcimônia).

Para a criação de imagens digitais com fins de preservação digital, deve-se utilizar o formato TIFF em sua versão base, sem compressão. Deve-se, ainda, testar os arquivos produzidos nos *softwares* de acesso que se planeja utilizar, pois um *software* pode ser incapaz de interpretar o TIFF gerado por outro.

Outra desvantagem do formato é a falta de suporte nativo que ele possui com relação aos navegadores de Internet mais populares (Microsoft Internet Explorer e Mozilla Firefox). Sendo assim, caso seja necessária a exibição dos documentos via WEB, estes deverão ser convertidos para um formato compatível como JPEG.

5.1.2. Joint Photographic Experts Group – JPEG

O JPEG é um exemplo de uma técnica de compressão que se tornou sinônimo de formato de imagem digital. Inicialmente, o *Joint Photographic Experts Group* definiu a técnica de compressão JPEG e a embutiu na especificação do formato JFIF (*JPEG File Interchange Format*), mas o nome JPEG logo se tornou referência na descrição de imagens que utilizam a compressão [TASI 02a].

A compressão JPEG é com perdas (*lossy compression*), ou seja, informações das imagens são eliminadas de forma permanente. Baseando-se em características do sistema visual humano, tal como a maior facilidade na distinção de brilho do que de cores, a técnica é capaz de reduzir a imagem original a até 1% do seu tamanho original, embora reduções a menos de 10% do original gerem distorções óbvias na imagem [TASI 03a].

Em consequência da grande redução proporcionada, o JPEG é o formato ideal para situações quando o tamanho é um fator mais crucial que a qualidade, por exemplo:

- Imagens destinadas à exibição via Internet, geralmente de qualidade inferior e conseqüentemente de menor tamanho. O formato possui suporte nativo nos principais navegadores (Internet Explorer e Mozilla Firefox).

Imagens de página geradas a partir de manuscritos ou outros documentos textuais em que o interesse está no conteúdo do texto e no tamanho final dos arquivos digitais, não em detalhes gráficos do documento. Entretanto, se existe a intenção de tratar o texto posteriormente com OCR para recuperar os caracteres, não se deve utilizar uma taxa de compressão muito agressiva.

O JPEG é um padrão aprovado pela ISO/IEC³ e, assim como o TIFF, tornou-se também um padrão industrial *de facto*. O formato possui suporte em praticamente todos os principais aplicativos de processamento de imagens digitais. Além disso, sua popularidade tende a aumentar devido à disseminação de câmeras fotográficas digitais e à distribuição/publicação de imagens na Internet (blogs, fotoblogs etc.).

5.2. Formatos para Documentos Textuais

Documentos textuais são componentes importantes em qualquer projeto de preservação digital. Seja através do armazenamento de metadados das imagens digitais, seja na preservação de textos que possuam valor histórico, tais como discursos, manuscritos, dentre outros.

É importante estar atento aos propósitos e limitações dos vários formatos disponíveis, de forma a ter certeza de que eles atendam aos requisitos e objetivos desejados sem limitar opções e mudanças futuras. Novamente salienta-se a preferência por formatos não-proprietários, pois, embora os formatos proprietários possuam boas ferramentas para geração de material publicável, no momento da transferência de dados para outro formato existe o risco de perda de formatação e até mesmo de informação do documento.

São três as maneiras utilizadas para o armazenamento de um documento textual:

- Imagem de página (*page image*);
- Texto puro (*plain text*);
- Uma forma híbrida contendo textos e imagens.

³ International Organization for Standardizations/International Electrotechnical Commission.

Uma imagem de página, uma imagem digital de um documento textual, capturada por um scanner ou câmera digital e apresentada como uma sequência de *pixels*, deve ser armazenada num dos formatos apresentados na Seção 5.1.

O texto puro corresponde ao documento armazenado como uma sequência de caracteres. Em sua versão mais simples, o texto puro é composto apenas por caracteres ASCII ou Unicode não possuindo qualquer informação referente à estrutura do documento. Versões “incrementadas” fazem uso de marcações de diversos tipos para descrever a estrutura e formatação do documento. Em ambos os casos, entretanto, é possível a pesquisa, no documento, por palavras e frases específicas, característica ainda não disponível nas imagens de página.

Já a abordagem híbrida incorpora as principais vantagens de ambos os métodos descritos acima: apresenta um documento formatado e com imagens, assim como uma imagem de página, e permite a pesquisa por palavras e frases, assim como um texto puro.

Apresenta-se nas seções que seguem detalhes sobre as codificações ASCII e Unicode, o formato XML, que faz uso de marcações, e um exemplo da abordagem híbrida: o formato PDF.

5.2.1. Texto Puro

A forma mais simples de se representar e armazenar um documento textual é através do uso de caracteres ASCII ou Unicode, o chamado texto puro. Ao contrário de uma imagem de página; tal documento pode ser pesquisado por palavras ou frases, mas não contém nenhuma informação sobre a aparência original e a estrutura do documento. O texto puro deve ser utilizado quando houver a necessidade de armazenar apenas o conteúdo de um documento textual, ou seja, quando a informação sobre a forma de apresentação do documento for considerada irrelevante. O uso dessa estratégia gera um certo grau de independência com relação à plataforma utilizada, uma vez que os dados podem ser exportados para praticamente qualquer outro formato.

A falta de estrutura característica em um texto puro, entretanto, gera o inconveniente de o usuário ter que conhecer bem a informação ali armazenada. Só assim é possível a correta interpretação do seu conteúdo, seja por um ser humano ou de forma automatizada pelo computador.

Como dito no início desta seção, são duas as principais formas de representação de um texto puro: caracteres que utilizam a codificação ASCII e caracteres que utilizam a codificação Unicode.

5.2.1.1 ASCII

Proposto em 1963 pelo *American National Standards Institute* (ANSI), o ASCII, acrônimo para *American Standard Code for Information Interchange*, foi desenvolvido para facilitar a compatibilidade entre os variados sistemas de processamento de dados então existentes. O ASCII utiliza os 128 números possíveis de se representar com 7 bits — de 0 até 127 — para codificar letras, números, sinais de pontuação e alguns caracteres especiais. Existem algumas extensões para o conjunto de caracteres ASCII que designaram os dígitos entre 128 e 255 para a representação de caracteres gráficos, símbolos matemáticos e caracteres não presentes na língua inglesa [GREEN 03].

5.2.1.2 Unicode

O formato foi desenvolvido com o objetivo de reparar falhas presentes em sistemas de codificação mais antigos, como o ASCII, por exemplo, em que um mesmo código pode ser utilizado por dois caracteres ou códigos diferentes representavam o mesmo caractere [GREEN 03].

O Unicode utiliza um código único para cada caractere, e este código tem estabilidade garantida, ou seja, não sofrerá mudanças em versões futuras do padrão. O suporte multiplataforma, multilíngua, e a codificação de símbolos técnicos e matemáticos são outras características marcantes do Unicode.

5.2.2. eXtensible Markup Language – XML

O XML é uma versão etiquetada de um texto puro (ver Seção 5.2.1). XML é um exemplo de linguagem de marcação (*markup language*) assim como o HTML usado na construção de páginas WEB. O HTML, entretanto, descreve o conteúdo apenas em termos de como ele deve ser exibido. Já o XML descreve o conteúdo em termos do que os dados representam, separando desta forma estrutura da apresentação do documento.

Documentos XML podem ser utilizados como um formato base a partir do qual podem ser geradas versões do documento para propósitos específicos:

- Uma versão para impressão;
- Uma página HTML para exibição na WEB;
- Uma versão Braille para deficientes visuais, dentre outras.

Outra característica da linguagem encontra-se na sua extensibilidade. As marcações (também chamadas *tags*, ou etiquetas) no XML não são fixas nem predefinidas, devendo o usuário, portanto, desenvolver um conjunto de marcações

adequado a suas necessidades. *Schemas* ou DTD's⁴ são as formas utilizadas pelo padrão para a definição do significado destes conjuntos de marcações. Para que um documento XML tenha seu significado realmente compreendido, é preciso um acordo com relação ao significado das *tags*.

Sendo um padrão industrial internacional, o XML tem grande suporte de empresas de *software* e de entidades padronizadoras, principalmente o W3C, que garantem a longevidade do formato. O XML é ainda independente de *hardware* e *software*, portanto, ideal para a criação de documentos textuais de preservação.

5.2.3. Portable Document Format – PDF

O PDF é um formato de arquivos digitais desenvolvido pela *Adobe* em 1992. Embora seja proprietário, o formato possui especificação pública possibilitando a qualquer pessoa ou empresa desenvolver aplicativos que utilizem a tecnologia.

O formato PDF é capaz de armazenar texto e imagens de forma estruturada, o que o torna uma opção atraente quando o objetivo é a preservação não só do conteúdo do documento, mas também de sua forma de apresentação.

Além de preservar a estrutura do documento, um arquivo PDF permite, ainda, a busca por palavras e sentenças no texto. Esta característica é interessante, pois permite que arquivos PDF que estejam disponíveis na WEB, sejam indexados por ferramentas de busca como, por exemplo, o *Google*. É possível, ainda, encapsular em um arquivo PDF metadados utilizando a tecnologia XML [ADOBE 02].

O PDF, assim como o TIFF, tornou-se um padrão industrial *de facto*, especialmente com o advento da Internet. A existência de *plug-ins* para os principais navegadores fez do formato uma forma conveniente de troca e visualização de documentos via WEB.

A independência de plataforma é outra vantagem do formato. Um arquivo PDF é visualizado da mesma maneira em sistemas *Microsoft Windows*, *Unix*, *Linux* ou *Mac OS* [ADOBE 02]. A própria *Adobe* fornece gratuitamente o *software Acrobat Reader* para a visualização de documentos PDF, além do *software* não-gratuito *Adobe Acrobat* para a produção de documentos PDF. Existem, entretanto, outros *softwares* que produzem PDF, inclusive, gratuitos.

⁴ *Document Type Definition*, ou seja, Definição de Tipo de Documento.

5.3. Padronização de Formatos

A diversidade dos formatos de arquivos existentes torna a preservação dos registros digitais complexa e onerosa. A padronização consiste na adoção de um formato único para cada tipo de registro digital (textos, imagens, vídeos), escolhe-se um formato para uma determinada classe de registros, e todos os exemplares daquele tipo são convertidos ao formato selecionado.

Os ganhos com a padronização são óbvios: utilizando-se um único formato, tarefas como refrescamento e migração de registros são simplificadas e por consequência têm seus custos reduzidos. O acompanhamento da evolução se concentra em poucos formatos, permitindo uma facilidade maior na identificação de novas tendências tecnológicas, dentre outras.

5.4. Tabela Comparativa

Formato	Aplicação	Características
TIFF	Imagens digitais de preservação.	• Independente de plataforma; • Formato proprietário, mas com especificação aberta; • Padrão industrial <i>de facto</i>; • Não utiliza compressão de dados em sua forma básica.
JPEG	Imagens para exibição na WEB. Imagens de página.	• Independente de plataforma; • Formato de especificação aberta; • Padrão internacional ISO/IEC; • Padrão industrial <i>de facto</i> (formato da WEB); • Utiliza compressão de dados com perda.
Texto Puro	Documentos textuais simples (sem estruturação e <i>layout</i>).	• Independentes de plataforma; • Padrões internacionais: ANSI/ASCII e ISO/Unicode; • Permite pesquisas por elementos textuais (frases e palavras).
XML	Documentos textuais de preservação.	• Independente de plataforma; • Recomendado e mantido pelo W3C; • Padrão industrial <i>de facto</i>; • Formato autocontido, o documento "sabe" o que armazena; • Permite pesquisas por elementos textuais (frases e palavras).
PDF	Documentos híbridos, textos e imagens.	• Independente de plataforma; • Formato proprietário, mas com especificação aberta; • Padrão industrial <i>de facto</i>; • Preserva o <i>layout</i> do documento; • Permite pesquisas por elementos textuais (frases e palavras).

Tabela 4: Principais formatos para arquivos digitais de preservação.

6. MÍDIAS DE ARMAZENAMENTO

A produção de imagens digitais apenas não é capaz de garantir longevidade a acervos digitais. Assim como acervos analógicos, os acervos digitais devem ser armazenados em suportes que garantam o mínimo de proteção contra danos.

Desde seu surgimento, os vários suportes para armazenamento digital têm apresentado um impressionante avanço tecnológico resultando em melhorias na capacidade, velocidade e confiabilidade destes dispositivos. Estas melhorias têm sido impulsionadas pelo aumento da capacidade de armazenamento de dados em uma mesma área (a chamada densidade de armazenamento). A maior densidade de armazenamento, entretanto, torna mais graves os problemas associados a esses suportes, pois pequenos defeitos tornam grandes volumes de informação indisponíveis.

"Os arquivistas estão tradicionalmente acostumados a lidar com suportes cuja durabilidade se mede em séculos, e não em anos. Se a disseminação do papel de polpa de madeira química — cuja constituição acídica o leva a uma lenta autodestruição ao longo de 100 ou 150 anos — já foi motivo de grave preocupação, as mídias digitais, com suas curtíssimas vidas, por vezes de menos de uma década, representam um desafio que não se poderia conceber no passado." [VALLE 03]

Lidando com suportes para armazenamento digital, é importante possuir algum conhecimento sobre a capacidade dos mesmos. A unidade fundamental do armazenamento digital é o *bit*. A partir da aglutinação de vários *bits* é possível obter-se grandes volumes de armazenamento. A listagem a seguir apresenta algumas equivalências entre medidas utilizadas no restante do texto.

8 bits	=	1 byte
1024 bytes	=	1 Kbyte (Kilobyte)
1024 Kbytes	=	1 Mbyte (Megabyte)
1024 Mbytes	=	1 Gbyte (Gigabyte)
1024 Gbytes	=	1 Tbyte (Terabyte)

Nas seções que seguem discorre-se sobre os três principais grupos de mídias de armazenamento utilizados na preservação de acervos digitais: discos rígidos,

fitas magnéticas e discos óticos. Apresentam-se ainda algumas recomendações sobre o ambiente de armazenamento das mídias.

6.1. Discos Rígidos

Discos rígidos, também conhecidos como HDs,⁵ são dispositivos eletromecânicos constituídos de um ou mais discos, cobertos por material magnetizável. Estes discos são acoplados a um eixo onde giram a altas velocidades (entre 5.000 e 10.000 RPM). Ao contrário dos demais suportes aqui apresentados, os discos rígidos possuem a mídia de dados e o mecanismo de leitura/escrita agrupados num mesmo equipamento (Figura 1) [VALLE 03].

Este tipo de suporte é indicado para o armazenamento de arquivos de uso corrente, imagens para acesso WEB, banco de dados *on-line*, código executável de aplicativos e do sistema operacional.

Avanços tecnológicos têm permitido aos fabricantes reduzir os discos rígidos em tamanho e consumo de energia e expandir sua capacidade de armazenamento. Encontram-se HDs de alta capacidade (superior a 100 GB) com custo de armazenamento inferior a 4 reais/Gbyte.

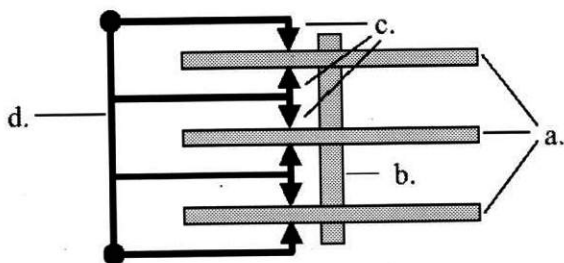


Figura 1: Perfil esquemático de um disco rígido, mostrando os a) discos montados sobre o b) eixo, as c) cabeças de leitura/gravação e o d) braço de acionamento das cabeças.

FONTE – [VALLE 03]

A sofisticação desses equipamentos tem seu preço: discos rígidos são equipamentos delicados devendo, portanto, ser manuseados com muito cuidado. Evite qualquer tipo de choque ou transporte com o equipamento ligado: o disco é

⁵ Abreviação do termo inglês *hard drive*.

mais sujeito a danos mecânicos nestas situações. As falhas mais comuns são o choque das cabeças de leitura/gravação com a superfície de um dos discos, mau funcionamento de motores e problemas de lubrificação. A Tabela 5 apresenta situações a executar ou evitar no manuseio e conservação de HDs.

Evite	Nunca	Sempre
Ambientes com fornecimento de energia irregular.	Aplique choques mecânicos no disco.	Execute testes de integridade do disco periodicamente.
Transporte desnecessário do equipamento.	Desligue o disco durante a gravação de dados.	Desfragmente o disco periodicamente.
	Manuseie o disco ou transporte o computador, estando este ligado.	

Tabela 5: Recomendações de preservação para discos rígidos.

FONTE – [GREEN 03]

Não se espera dos discos rígidos uma vida útil superior a 4 ou 5 anos, requerendo-se, portanto, um acompanhamento das datas de instalação, e um plano de substituição progressivo.

O uso efetivo destes suportes está intimamente relacionado com os cuidados dispensados a eles. Seguindo as recomendações acima, pode-se esperar destes equipamentos uma operação contínua e confiável, durante sua curta vida útil.

6.2. Fitas Magnéticas

Esta classe de suportes digitais representa o amplo espectro de mídias que utilizam uma camada de material magnético passível de gravação sobre um substrato flexível — a fita propriamente dita [VALLE 03].

Fitas magnéticas são apresentadas em vários formatos, sendo os mais populares o cassete e o cartucho. Existe ainda uma terceira forma de apresentação de fitas magnéticas: em rolo. Este formato, entretanto, é antigo e está em desuso. Fitas em cassete embutem um rolo doador e um rolo receptor em um único invólucro e são hoje em dia as mais difundidas. Cartuchos possuem um único rolo: a fita se apresenta ou como um laço sem fim ou com uma guia inicial que é adaptada a um segundo rolo embutido no dispositivo de leitura e gravação.

Dentre as várias opções de fitas magnéticas existentes no mercado, merecem atenção a DDS,⁶ no formato cassette, e a DLT⁷ representando o formato cartucho (Figura 2).

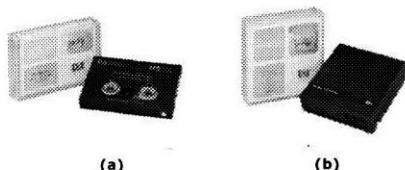


Figura 2: Exemplos de fitas magnéticas em a) cassette (DDS-4) e b) cartucho (DLT Type-IV).

A fita DDS, um cassette de 4mm baseado na tecnologia DAT,⁸ apresenta, em sua versão mais recente, a DDS-4, capacidade de 20 GB podendo alcançar 40 GB quando utilizada em modo comprimido [VALLE 03]. Já a fita DLT, de patente da *Quantum Corporation*, também em sua quarta versão (DLT Type-IV), possui capacidades variando entre 20 e 80 GB. Novas versões desta tecnologia prometem volumes superiores a 1 TB [DLT TAPE].

Um aspecto negativo com relação a fitas magnéticas está no armazenamento sequencial dos dados. Esta característica pode resultar em acessos lentos para operações de leitura/escrita, principalmente as realizadas no final da fita.

Os cuidados com este suporte ocorrem em duas frentes, com os leitores da mídia (popularmente conhecidos como *drives*) e com as fitas propriamente ditas. Os leitores, dispositivos mecânicos, tendem a acumular poeira que pode ser passada à fita, reduzindo sua expectativa de vida. A principal preocupação, entretanto, deve ser com o mecanismo de leitura/escrita. O contato entra a cabeça e a fita gera um desgaste natural que pode ser acentuado quando o equipamento está desregulado. Fitas próprias para a limpeza dos *drives* devem ser utilizadas periodicamente. Com relação às fitas, a Tabela 6 apresenta algumas recomendações úteis para a conservação das mesmas.

⁶ Do inglês *Digital Data Storage*.

⁷ Do inglês *Digital Linear Tape*.

⁸ Do inglês *Digital Audio Tape*.

Evite	Nunca	Sempre
Armazenar as fitas próximas a campos magnéticos.	Empilhe as fitas horizontalmente.	Mantenha as fitas em suas caixas de proteção quando não estiverem em uso.
Transportar as fitas desprotegidas.	Utilize etiquetas fora das recomendações do fabricante nem afixe-as fora das áreas reservadas.	Transporte as fitas em suas caixas.
Expor as fitas à variações bruscas de temperatura ou umidade relativa.	Toque na superfície da fita.	Armazene as fitas de forma apropriada, em posição vertical.

Tabela 6: Recomendações de preservação para fitas magnéticas.

FONTE – [GREEN 03]

Uma atenção adicional deve ser considerada se a área de armazenagem permanente das fitas tiver condições atmosféricas (temperatura e umidade) significativamente diversas das da sala do computador. Neste caso, a fita deve ficar em repouso durante várias horas, aclimatando-se, antes de ser inserida no dispositivo.

Fitas são mídias longevas, mas frágeis, e não admitem limpeza deficiente nem negligência nas operações. Fumo, bebida e comida devem ser estritamente proibidos nos ambientes onde elas são manipuladas, e todos os usuários devem receber treinamento adequado antes de acessá-las.

6.3. Discos Óticos

A tecnologia de discos óticos é a mais recente dentre os suportes apresentados. Ganhou destaque em meados da década de 80 com a popularização dos discos compactos,⁹ ou CDs, como um suporte para música digital. Com o sucesso do CD, outros tipos de dados começaram a utilizar a tecnologia, produzindo variantes como CD-ROM, ou disco compacto de memória apenas para leitura,¹⁰ para uso em computadores.

A disseminação dos discos compactos como suporte para armazenamento digital, entretanto, só foi possível com o desenvolvimento de mídias para gravação

⁹ Do inglês *Compact Disc*.

¹⁰ Do inglês *Compact Disc Read-Only Memory*.

em pequena escala — o CD-R e o CD-RW. A gravação de CDs de áudio e CD-ROMs faz uso de processos mecânicos viáveis apenas em escala industrial, daí a necessidade de novos formatos. O CD-R é um disco compacto que permite uma única gravação da mídia enquanto que o CD-RW suporta regravação.

Atualmente, um disco compacto, seja ele um CD-ROM, CD-R ou CD-RW, possui capacidade de armazenamento equivalente a 700 MB. A baixa capacidade de armazenamento é uma das principais desvantagens dos discos compactos, principalmente quando comparados aos modernos discos rígidos e fitas magnéticas cuja capacidade é medida em dezenas ou até centenas de gigabytes.

Para suprir esta deficiência, a indústria apresentou o disco digital versátil,¹¹ ou DVD. Externamente um DVD é semelhante a um CD, mas os avanços do primeiro em relação ao segundo são enormes [VALLE 03]:

- Maior densidade de dados;
- Codificação mais eficiente;
- Possibilidade de duas camadas de dados em cada um dos lados do disco;
- Possibilidade de gravação em ambos os lados do disco.

Estas melhorias resultaram em uma capacidade de armazenamento considerável. A versão mais simples de um DVD armazena até 4.7 GB de dados e quando na versão mais poderosa (duas camadas, dupla face) sua capacidade atinge os 17 GB.

Como aconteceu com o CD, o DVD foi inicialmente desenvolvido visando à produção industrial. A produção em menor escala, entretanto, não tardou em ganhar versões da tecnologia dando início a uma gama de formatos que hoje representam uma fonte de confusão para os usuários [KENNEY 03b]. Existem cinco formatos disputando o mercado doméstico — DVD-R, DVD-RW, DVD+R, DVD+RW e DVD-RAM — e a inexistência de um padrão único levou à incompatibilidade entre os leitores, gravadores e mídias. Felizmente, o advento de *drives* com suporte a múltiplos formatos, os chamados multi-DVDs, mitigou os problemas de incompatibilidade, finalmente tornando o uso e intercâmbio dessas mídias conveniente.¹²

A durabilidade de mídias óticas tem sido ponto de polêmica, principalmente entre mídias graváveis. Alguns fabricantes garantem durabilidade de até 100

¹¹ Do inglês *Digital Versatile Disc*.

¹² Deve-se ter em mente que mesmo leitores *multi-DVD* não apresentam total compatibilidade entre formatos.

anos,¹³ enquanto registram-se casos de deterioração de mídias em menos de um ano. O que é certo é que mídias apresentam enormes variações de durabilidade de acordo com o fabricante, ou seja, marcas muito baratas ou desconhecidas devem ser evitadas.

Não obstante à qualidade das mídias, discos óticos estão sujeitos a muitos riscos. Arranhões e resíduos diversos (gordura da mão, poeira etc.) não apenas podem causar perdas de dados como também são fontes de contaminação que irão gradualmente destruir a proteção dos discos tornando estes completamente inutilizáveis.

Fazendo uso de mídias de melhor qualidade e observando-se as recomendações presentes na Tabela 7, é certa uma longevidade de pelo menos uma década para suportes óticos e — se os ensaios de envelhecimento acelerado forem corretos — talvez um século.

Evite	Nunca	Sempre
Danificar a superfície superior e inferior dos discos, bem como sua lateral.	Adicione ou fixe nada à superfície dos discos.	Mantenha os discos em suas caixas de proteção ou capas quando não estiverem em uso.
Fricção ou contato com superfícies que possam engordurar o disco (e.g. mão humanas).	Escreva em parte alguma do disco a não ser na área reservada para tal.	Utilizando capas, dê preferência àquelas de algodão, com qualidade arquivística.
Expor os discos, diretamente, à luz solar.		Use luvas ao manusear os discos mestres.

Tabela 7: Recomendações de preservação para discos óticos.

FONTE – [GREEN 03]

Para preservação em longo prazo, as caixas são superiores às capas, envelopes ou luvas. As caixas são concebidas de forma a não haver contato físico com a superfície do disco, ao contrário do que ocorre com capas e luvas. Capas apertadas podem sujeitar o disco a deformações, em especial na presença de variações de temperatura.

¹³ Por serem dispositivos recentes, os testes de durabilidade dos discos óticos são resultados de ensaios de envelhecimento acelerado e não da observação direta de situações reais. [VALLE 03]

6.4. Comparação entre as Mídias de Armazenamento

A Tabela 8 apresenta uma comparação entre as diversas mídias encontradas no mercado. Leva-se em conta o custo do dispositivo de leitura/escrita, o custo unitário da mídia e sua capacidade, e a partir destes dados obtém-se a quantidade de reais necessária por gigabyte armazenado.

Mídia	Custo do drive ^a	Custo unitário da Mídia ^a	Capacidade (GB)	Reais/Gbyte
Disco rígido	R\$ 400,00 – R\$ 1.000,00	—	120 - 250	R\$ 3,00 – R\$ 4,00
DDS-4	R\$ 3.500,00 – R\$ 5.000,00	R\$ 40,00 – R\$ 80,00	20	R\$ 2,00 – R\$ 4,00
DLT Type-IV	R\$ 5.000,00 – R\$ 15.000,00	R\$ 140,00 – R\$ 180,00	40	R\$ 3,50 – R\$ 4,50
CD-R/RW	R\$ 200,00	R\$ 1,00 – R\$ 5,00	0,7	R\$ 1,50 – R\$ 7,00
DVD-RAM	R\$ 500,00	R\$ 40,00 – R\$ 100,00	2,6 – 9,4	R\$ 10,50 – R\$ 15,50
DVD-R/RW*	R\$ 500,00	R\$ 9,00 – R\$ 55,00	4,7 – 9,4	R\$ 1,90 – R\$ 6,00
DVD+R/+RW*	R\$ 500,00	R\$ 11,00 – R\$ 16,00	4,7	R\$ 2,50 – R\$ 3,50

Tabela 8: Quadro comparativo entre mídias de armazenamento.

^a Os custos associados às mídias são valores aproximados; faixas de valores correspondem aos preços máximos e mínimos encontrados.

* Discos rígidos não têm custo unitário associado, pois, como dito anteriormente, possuem a mídia e o mecanismo de leitura/escrita num mesmo dispositivo.

6.5. Ambiente de Armazenamento

Em nada resultam os cuidados com os suportes digitais recomendados nas seções anteriores se os mesmos são armazenados com desleixo. Sendo assim, o ambiente de armazenamento influi consideravelmente na qualidade e durabilidade dos acervos digitais.

Todos os suportes devem ser armazenados em ambientes que ofereçam proteção contra a incidência direta de luz solar e luz ultravioleta, e que estejam livres de poluentes atmosféricos.

Áreas de armazenamentos de mídias magnéticas não devem possuir grandes equipamentos elétricos ou qualquer equipamento capaz de gerar campos magnéticos. Estes ambientes ainda não devem apresentar campos magnéticos ocasionais superiores a 4000 Amperes/metro.

A Tabela 9 e a Tabela 10 apresentam, respectivamente, condições ambientais recomendadas para a preservação de curto e de longo prazo.

Mídia	Temperatura	Umidade Relativa do Ar
Discos Rígidos	5-55°C	5-95%
Digital Data Storage (DDS)	5-45°C	20-80%
Digital Linear Tape (DLT)	16-32°C	20-80%
Outras fitas magnéticas	10-45°C	20-80%
CD-ROM/R/RW	10-50°C	10-80%
DVD-ROM/R/+R/RAM/RW/+RW	10-50°C	10-80%
Coleções mistas	16-32°C	20-80%

Tabela 9: Condições ambientais para armazenamento em curto prazo.

FONTE – [BROWN 03b]

Mídia	Temperatura	Umidade Relativa do Ar
Discos Rígidos	18-22°C	35-45%
Digital Data Storage (DDS)	5-32°C	20-60%
Digital Linear Tape (DLT)	18-26°C	40-60%
Outras fitas magnéticas	18-22°C	35-45%
CD-ROM/R/RW	18-22°C	35-45%
DVD-ROM/R/+R/RAM/RW/+RW	18-22°C	35-45%
Coleções mistas	18-22°C	35-45%

Tabela 10: Condições ambientais para armazenamento em longo prazo.

FONTE – [BROWN 03b]

7. ESTRATÉGIAS DE PRESERVAÇÃO DIGITAL

7.1. Refrescamento

O refrescamento é a etapa da preservação digital relacionada com a deterioração dos suportes e tem como objetivo evitar que a degradação física e a obsolescência tecnológica levem a uma perda parcial ou total da informação digital neles contida.

Este procedimento consiste na simples transferência dos dados de um suporte para outro, não envolvendo, portanto, alteração nos formatos e padrões dos arquivos digitais [KENNEY 03a]. O refrescamento é realizado tanto para a substituição de uma mídia obsoleta, substituindo, por exemplo, disquetes por DVDs, como para a transferência de dados de uma mídia degradada para outra em melhores condições.

O refrescamento não implica em qualquer mudança no formato dos arquivos digitais, estes são copiados *bit-a-bit* entre os suportes durante o procedimento.

O refrescamento pode ser considerado o procedimento mais simples de um programa de preservação digital, no entanto, definir o prazo para sua realização requer planejamento e acompanhamento sistemático da integridade dos suportes e da evolução da tecnologia de armazenamento.

A verificação da integridade das mídias garante que o refrescamento ocorra no momento correto, pois atrasos aumentam o risco de perda de dados. Suportes de qualidade, manuseio cuidadoso, controle da umidade e temperatura do ambiente de guarda são fatores que influenciam na durabilidade das mídias e podem retardar o refrescamento (ver Capítulo 6).

O acompanhamento da tecnologia de armazenamento permite a identificação de tecnologias promissoras, possibilitando ao arquivista adiantar-se à obsolescência dos suportes por ele utilizados.

Um refrescamento bem sucedido envolve cuidados habituais em trabalhos de reformatação: controle da retirada dos materiais de seus depósitos permanentes, separação cuidadosa entre mídias que já foram refrescadas e as que ainda vão ser, identificação correta de novos materiais com etiquetas, controle de qualidade na gravação, verificação da integridade dos registros transferidos para nova mídia.

Em muitas situações, o refrescamento irá ocorrer entre suportes de tecnologias diferentes, geralmente, tendo o novo suporte maior capacidade. Assim, o novo suporte poderá aglutinar o conteúdo de várias unidades do suporte obsoleto: um DVD, por exemplo, é capaz de reunir informações de até 7 CDs.

O impacto disso sobre a preservação está na indexação dos dados, essencial para que se possa recuperá-los depois. Uma determinada instituição poderia ter os arquivos "0001.TIF" e "0002.TIF" referentes às páginas do documento X armazenados no CD "C-001", e outros arquivos "0001.TIF" e "0002.TIF" referentes às páginas do documento Y armazenados no CD "C-005". No processo de reformatação para DVD-R, os CDs "C-001" até "C-007" são copiados para o DVD-R "DVD-001". O que deve ser feito? Separar os arquivos no DVD-R em pastas distintas é uma possibilidade, renomear os arquivos para "001-0001.TIF" e "005-0001.TIF" etc., é uma outra. Em qualquer caso, será necessário alterar as referências ao arquivo original, de forma que elas reflitam a nova localização do arquivo, e/ou, criar uma tabela de equivalência entre as referências antigas e as novas [VALLE 03].

7.2. Migração

A migração, assim como o refrescamento, é uma operação periódica, e seu propósito é preservar a integridade dos dados digitais, dando a garantia de que estes tenham condições de ser visualizados e utilizados pelos usuários. Entretanto, ao contrário do refrescamento, onde os registros não sofrem modificações, a migração acarreta a mudança dos registros para formatos suportados pelas novas configurações de *hardware/software*. Assim, a migração pode ser definida como um conjunto de procedimentos organizados desenvolvidos para a transferência periódica de materiais digitais de uma configuração de *hardware/software* para outra, ou de uma geração tecnológica para a geração subsequente.

São exemplos de migração a transferência de um arquivo *Wordstar* para *Microsoft Word*, a conversão de uma imagem digital do formato TIFF para o formato JPEG, assim como a atualização de versões do *Microsoft Excel 5.0* para *Microsoft Excel 2003*.

Como qualquer processo de tradução, a migração sujeita os dados a uma certa distorção de significado, além de ser uma operação mais complexa e custosa que o refrescamento [VALLE 03]. A complexidade varia de acordo com o tipo de dado migrado (texto, imagem, vídeo, som) e do resultado desejado. Na migração de um texto, por exemplo, é preciso responder à questão: o que se deseja preservar,

somente a informação contida no documento ou, além disso, o *layout* do mesmo? Respostas a perguntas como esta norteiam todo o processo, desde a seleção do novo formato de arquivo passando por testes de verificação, controle de qualidade, dentre outros.

Experiências demonstram que migrações, aparentemente simples, podem produzir alterações visuais nos documentos: por exemplo, na migração para versões mais novas de documentos no formato PDF, em algumas ocasiões, ocorreram deslocamentos da última palavra de uma linha do texto original para linha seguinte. Em outros documentos textuais observou-se a perda de formatação de estilos de fontes (negrito e itálico) nos arquivos convertidos.

A correta realização da migração é precedida de análises e cuidados sobre a realização desta estratégia de preservação:

- Avaliar criteriosamente os formatos digitais com o objetivo de utilizar aquele que mais se adapte aos objetivos da migração (ver Capítulo 5).
- Realizar testes preliminares antes da completa execução da migração, verificando assim se o procedimento está de acordo com as expectativas.
- Estabelecer um controle de qualidade dos resultados da migração.
- Gerar metadados com informações sobre o processo visando à documentação das alterações pelas quais o registro venha a passar durante sua existência.

Seguindo estas recomendações têm-se, além de uma migração mais segura, uma redução nos custos envolvidos na operação. É importante salientar, entretanto, que os procedimentos de migração devem ser realizados antes que os formatos utilizados tornem-se obsoletos.

7.2.1. Relação entre Refrescamento e Migração

Migração é um conceito mais amplo e rico que o refrescamento por identificar uma gama de opções para a preservação digital. O objetivo da migração é preservar a integridade digital dos registros para manter a capacidade dos mesmos de serem recuperados, visualizados e utilizados em face à constante evolução tecnológica. A migração inclui o refrescamento como um meio de preservação digital, entretanto difere dele na medida em que não realiza uma cópia digital exata, mas procura manter a integridade do registro enquanto o traduz para um formato compatível com a nova geração tecnológica.

7.3. Emulação

Um terceiro procedimento de preservação digital é a emulação. Na emulação, preocupa-se com a preservação intelectual da plataforma computacional (*hardware/software*) na qual os suportes digitais estão inseridos. Este procedimento "procura manter documentos digitais em suas plataformas originais, enquanto cria ambientes operacionais que simulam, em novas plataformas, o comportamento daquelas que se tornaram obsoletas" [BESSER 00].

A emulação requer a criação de emuladores, programas tradutores de códigos e instruções de uma plataforma computacional para outra. Estes programas possibilitam a execução de aplicações obsoletas ou mesmo a visualização de documentos digitais ultrapassados.

Uma emulação de sucesso requer que informações sobre requisitos de *hardware* e *software* sejam preservadas de forma que a plataforma computacional possa ser recriada.

A criação dos programas emuladores, entretanto, caracteriza a principal deficiência do procedimento: produzir um emulador, seja de um *hardware* ou de um *software*, é uma tarefa complexa e custosa. A adoção da emulação como estratégia de preservação em longo prazo requer a criação de consórcios entre instituições de guarda (arquivos públicos e privados, museus e bibliotecas) para a realização dos aspectos técnicos necessários, bem como para redução dos custos envolvidos (especialmente os relacionados à propriedade intelectual) [KENNEY 03a].

Migração vs. Emulação

Existe, na literatura sobre preservação digital, um grande embate sobre qual é a melhor estratégia de preservação digital: **migração** ou **emulação**.

Alguns autores apontam a emulação como a única forma confiável de preservar a funcionalidade, integridade e os aspectos visuais dos registros digitais. Os críticos da migração apontam que sua incapacidade em preservar a apresentação visual dos registros e seu risco à integridade física dos dados e, ainda, que as sucessivas migrações podem levar à perda parcial ou total dos registros digitais.

A emulação, por sua vez, é duramente criticada porque o emulador ao simular o funcionamento de plataformas obsoletas não permite que o usuário se beneficie dos avanços tecnológicos como: novas funcionalidades de programas, maior capacidade de processamento de computadores, melhorias na resolução de monitores, dentre outros. Outro aspecto criticado deve-se à necessidade de o usuário do presente conhecer o funcionamento de plataformas antigas.

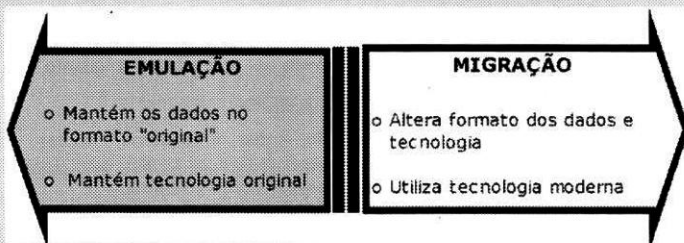


Figura 3: Emulação vs. Migração

8. CONCLUSÃO

Projetos de digitalização de acervos são, em sua maioria, voltados exclusivamente para a geração de cópias digitais, assim como para o desenvolvimento de sistemas de informação e pesquisa. No entanto, a fragilidade dos meios de armazenamento digitais e a obsolescência tecnológica representam um risco constante aos registros digitais e, conseqüentemente, aos investimentos em digitalização.

Projetos de digitalização somente podem ser justificados mediante a garantia da preservação dos registros digitais ao longo do tempo. O mundo digital exige que as ações de preservação tenham início no ato da criação do registro e, principalmente, devido à rápida evolução tecnológica, prossigam em todo seu ciclo de vida.

Projetos de digitalização não se encerram com a disponibilização das imagens digitais em um sistema de pesquisa, pois o compromisso com a preservação dos registros digitais requer continuidade. É preciso manter os registros em área de guarda apropriada, estabelecer intervalos para verificação da integridade dos dados, acompanhar o desenvolvimento tecnológico para assegurar que formatos e mídias não se tornem obsoletos, assim como realizar migrações periódicas - ou seja, as ações de preservação digital devem ser executadas enquanto houver interesse em manter os registros acessíveis.

Este *Caderno Técnico* apresentou, de forma introdutória, noções sobre preservação de acervos digitais que poderão auxiliar instituições interessadas na digitalização de seus acervos. Decisões sobre padrões e formatos de arquivos utilizados, definição e criação de metadados, cuidados com manuseio e armazenamento, etc., devem ser parte integrante dos projetos de digitalização, assim como do planejamento de ações, a médio e longo prazo, que garantam que os registros permaneçam acessíveis ao longo do tempo.

9. REFERÊNCIAS

- [ADOBE 02] ADOBE SYSTEMS INC. **PDF as a Standard for Archiving**. 2002.
Disponível na WWW: <<http://www.adobe.com/products/acrobat/pdfs/pdfarchiving.pdf>>. Acesso em 16 de agosto de 2004.
- [ADOBE 92] ADOBE SYSTEMS INC. **TIFF Specification – Revision 6.0**. 1992.
Disponível na WWW: <<https://partners.adobe.com/asn/developer/PDFS/TN/TIFF6.pdf>>. Acesso em 16 de agosto de 2004.
- [BESSER 00] BESSER, Howard. Digital Longevity. In: SITTS, Maxine (ed.). **Handbook for Digital Projects: a management tool for preservation and access**. Andover MA: Northeast Document Conservation Center, 2000. p.155-166. Acesso em 22 de maio de 2004.
- [BROWN 03a] BROWN, Adrian. **Digital Preservation Guidance Note 2: Selecting Storage Media for Long Term Preservation**. 2003. The National Archives, United Kingdom.
Disponível na WWW: <http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting_storage_media.pdf>. Acesso em 14 de junho de 2004.
- [BROWN 03b] BROWN, Adrian. **Digital Preservation Guidance Note 3: Care, handling and storage of removable media**. 2003. The National Archives, United Kingdom.
Disponível na WWW: <http://www.nationalarchives.gov.uk/preservation/advice/pdf/media_care.pdf>. Acesso em 14 de maio de 2004.
- [COLET 00] COLET, Linda S. *et al.* **Guides to Quality in Visual Resource Imaging**. 2000. Research Libraries Group, Digital Library Federation.
Disponível na WWW: <<http://www.rlg.org/visguides/>> Acesso em 17 de junho de 2004.
- [CPA 96] COMMISSION ON PRESERVATION AND ACCESS, the. **Preserving Digital Information: Report of the Task Force on Archiving of Digital Information**. 1996.
Disponível na WWW: <<http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>>. Acesso em 25 de maio de 2004.
- [CRUZ 02] CRUZ, Emília Barroso. **Memória governamental e utilização de documentos eletrônicos pela administração pública no Brasil: o caso da Secretaria de Estado da Fazenda de Minas Gerais**. Dissertação de Mestrado, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Brasil, 2002.

- [DLT TAPE] **DLT Tape.** Quantum Corporation.
Disponível na WWW: <<http://www.dlittape.com>> Acesso em 16 de setembro de 2004.
- [GRAY 04] GRAY, Joan. **Metadata.** 2004. Monash University Library. Monash University.
Disponível na WWW: <<http://www.lib.monash.edu.au/metadata/>>. Acesso em 21 de junho de 2004.
- [GREEN 03] GREEN, David L. *et al.* **The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials.** 1. ed. 2003. Humanities Advanced Technology and Information Institute, University of Glasgow and National Initiative for a Networked Cultural Heritage, New York University. Acesso em 20 de junho de 2004.
- [HOWELL 04] HOWELL, Allan G. **Preserving Digital Information:** Challenges and Solutions. 2004.
Disponível na WWW:
<http://www.alanhowell.com.au/papers/pdi_wkb.pdf> Acesso em 11 de agosto de 2004.
- [HUTT 04] HUTT, Arwen. **Understanding Metadata:** What is it really? The Tennessee Librarian, v. 54 n. 3, 2004. Knoxville.
Disponível na WWW:
<<http://www.lib.utk.edu/~tla/TL/v54n2/hutt.htm>>. Acesso em 14 de agosto de 2004.
- [KENNEY 01] KENNEY, Anne R., RIEGER, Oya Y., ENTLICH, Richard. **Preserving Cornell's Digital Image Collections:** Implementing an Archival Strategy. 2001 Cornell University Library, Cornell University. Acesso em 10 de julho de 2004.
- [KENNEY 03a] KENNEY, Anne R., *et al.* **Digital Preservation Management:** Implementing short-term strategies for long-term problems. 2003. Research Department, Cornell University Library, Cornell University.
Disponível na WWW:
<<http://www.library.cornell.edu/iris/tutorial/dpm/index.html>>. Acesso em 25 de agosto de 2004.
- [KENNEY 03b] KENNEY, Anne R., RIEGER, Oya Y., ENTLICH, Richard. **Moving Theory into Practice:** Digital Image Tutorial. 2003. Research Department, Cornell University Library, Cornell University.
Disponível na WWW: <<http://www.library.cornell.edu/preservation/tutorial/contents.html>>. Acesso em 15 de setembro de 2004.
- [NLA 03] NATIONAL LIBRARY OF AUSTRALIA. Preservation Metadata. 2003.
Disponível na WWW: <<http://www.nla.gov.au/padi/topics/32.html>>. Acesso em 27 de setembro de 2004.

- [ROSS 99] ROSS, Seamus e GOW, Ann. **Digital Archaeology: Rescuing Neglected and Damaged Data Resources**. 1999. Library Information Technology Centre South Bank University, London. 102 p. Acesso em 14 de agosto de 2004.
- [TASI 02a] TECHNICAL ADVISORY SERVICE FOR IMAGES. **Choosing a File Format**. 2002. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/creating/format.html>> Acesso em 10 de julho de 2004.
- [TASI 02b] TECHNICAL ADVISORY SERVICE FOR IMAGES. **An Introduction to Digital Image Preservation**. 2002. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/delivering/digpres.html>> Acesso em 13 de julho de 2004.
- [TASI 02c] TECHNICAL ADVISORY SERVICE FOR IMAGES. **Metadata and digital images**. 2002. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/delivering/metadata.html>> Acesso em 16 de setembro de 2004.
- [TASI 02d] TECHNICAL ADVISORY SERVICE FOR IMAGES. **Establishing a Digital Preservation Strategy**. 2002. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/delivering/digpres2.html>> Acesso em 11 de julho de 2004.
- [TASI 03a] TECHNICAL ADVISORY SERVICE FOR IMAGES. **File Formats and Compression**. 2003. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/creating/fformat.html>> Acesso em 12 de julho de 2004.
- [TASI 03b] TECHNICAL ADVISORY SERVICE FOR IMAGES. **New Digital Image File Formats**. 2003. Institute for Learning and Research Technology. University of Bristol.
Disponível na WWW:
<<http://www.tasi.ac.uk/advice/creating/newfile.html>> Acesso em 11 de julho de 2004.
- [VALLE 03] VALLE, Eduardo A. **Sistemas de informação multimídia na preservação de acervos permanentes**. Dissertação de Mestrado, Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Brasil, 2003.



Patrocínio



Apelo



Realização

